

EXHIBIT E

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 1

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

IN RE MATTER OF:)
RICHARD KADREY, et al.,)
Plaintiff,)
vs.) C.A. NO.:
META PLATFORMS, INC.,) 3:23-cv-03417-VC
Defendant.)
)

** HIGHLY CONFIDENTIAL - ATTORNEYS EYES ONLY **

VIDEOTAPED DEPOSITION OF AHMAD AL-DAHLE

Palo Alto, California

Thursday, October 3, 2024

stenographically Reported by:

HEATHER J. BAUTISTA, CSR, CRR, RPR, CLR
Realtime Systems Administrator
California CSR License #11600
Oregon CSR License #21-0005
Washington License #21009491
Nevada CCR License #980
Texas CSR License #10725

DIGITAL EVIDENCE GROUP
1730 M Street, NW, Suite 812
Washington, D.C. 20036
(202) 232-0646

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 57

1 foundation.

2 THE WITNESS: Not that I'm aware of, but
3 it's possible.4 Q. (By Mr. Young) Do you know who Shawn
5 Presser is?

6 A. Sorry. Say that again?

7 Q. Do you know who Shawn Presser is?

8 A. The name doesn't ring a bell.

9 Q. Do you know if Shawn Presser was the person
10 who originally created Books3?11 MR. GHAJAR: Objection. Lacks foundation.
12 Calls for speculation.13 MR. YOUNG: Counsel, I'll, again, request
14 that you limit your objections to form.15 THE WITNESS: The name doesn't ring a bell.
16 So I don't know what this person did or did not do.17 Q. (By Mr. Young) Do you know what the term
18 "shadow library" means?

19 A. No.

20 Q. Do you know if Meta ever obtained a copy of
21 the Books3 dataset?

22 MR. GHAJAR: Objection. Vague.

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 63

1 investigation into whether Plaintiffs' copyrighted
2 works appear in Books3?

3 MR. GHAJAR: Objection to the extent that
4 the answer would require the witness to reveal or
5 reflect information obtained from attorneys. I
6 would caution the witness not to reveal privileged
7 information.

8 THE WITNESS: Not outside of the legal
9 team.

10 Q. (By Mr. Young) Do you know what LibGen is?

11 A. I'm vaguely familiar with the term, but
12 very few specifics around it.

13 Q. When did you learn about LibGen? And I'm
14 not asking you to divulge any conversations you may
15 have had with attorneys.

16 A. It -- I was made aware of it recently as
17 part of this deposition.

18 Q. And if you're going to divulge any
19 conversations you may have had with attorneys, I'll
20 ask you to stop there.

21 A. Okay.

22 Q. Do you understand that LibGen is short for

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 64

1 Library Genesis, without divulging any conversations
2 you may have had with your counsel?

3 A. No.

4 Q. Are you aware if Meta used LibGen to train
5 any of the models it has published?

6 A. I'm not directly aware of that.

7 Q. So you are not aware, one way or the other,
8 if LibGen was included in the training corpora for
9 LLaMA 1; would that be fair to say?

10 A. It's important to remember I did not direct
11 LLaMA 1 development.

12 Q. So that would be no, because you did not
13 direct it?

14 A. I'm not deeply familiar with LLaMA 1's
15 development, because I did not oversee it.

16 Q. Would it be fair to say, based on your
17 testimony, you would not be aware, one way or the
18 other, if LibGen was used to train LLaMA 2?

19 A. I think it's fair to say two things: One,
20 I don't recall; and, two, just a reminder that when
21 we do make decisions on these sorts of datasets, at
22 least when I was overseeing LLaMA development, we

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 65

1 are very diligent about working with our --

2 (Stenographer clarification.)

3 THE WITNESS: -- with our cross -- with

4 cross-functional partners who represent the

5 different expert areas that we care about, like

6 policy, legal, et cetera.

7 Q. (By Mr. Young) Thank you.

8 What about LLaMA 3? Would you be aware,
9 one way or the other, if LibGen was used as any part
10 of the training corpora for LLaMA 3?

11 A. I don't recall specific discussions on it,
12 but, again, I suspect -- I speculate that we ran
13 processes around all of the decisions, key decisions
14 that we need to make during model development.

15 Q. And LLaMA 4 is currently in development at
16 Meta; right?

17 MR. GHAJAR: Counsel, this is a good time
18 for me to put on record that we'll designate this
19 transcript highly confidential, attorneys' eyes
20 only. Some of the discussion has gone into some
21 detail of business decisions at Meta, and it sounds
22 like you want to go into as-yet-developed products.

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 66

1 And so we'll so mark the transcript highly
2 confidential, attorneys' eyes only under the
3 protective order in the case.

4 THE WITNESS: Correct. We're working on
5 LLaMA 4.

6 Q. (By Mr. Young) And would you have an
7 awareness, one way or the other, if Meta is using
8 LibGen as part of the training corpora for LLaMA 4?

9 A. Important to remember when developing large
10 language models -- LLaMA 4 development is very
11 early, so it's not training. But it's important to
12 remember when developing these large language
13 models, data is only a simple component of what's
14 required to build them.

15 And actually, the underlying complexity and
16 capabilities in the model lie significantly in the
17 other components that we have to engineer and
18 design, such as training against large-scale
19 cluster. That usually spans tens of thousands of
20 GPUs. Dealing with, like, infant mortality, those
21 GPUs.

22 So majority of our development here is

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 67

1 focused on where we believe the greatest value
2 lives, which is the fundamental training mechanics
3 of the large language model.

4 Q. Thank you.

5 (Stenographer clarification.)

6 Q. (By Mr. Young) Thank you, Mr. Al-Dahle.

7 Mr. Al-Dahle, if I wanted to -- if --
8 who -- who could tell me whether or not LibGen was
9 used as part of the training for LLaMA 1?

10 A. I'm not sure I know. I have over a
11 thousand people, so it's quite complex. My
12 interactions with the team is usually via documents
13 and reviews.

14 Q. But you over- -- what about LLaMA 2? Would
15 you be able to tell me someone who would be able to
16 tell me whether or not LibGen was used to train
17 LLaMA 2?

18 MR. GHAJAR: Objection. Vague.

19 THE WITNESS: It's a big team, so I would
20 have to think about that, who would have the most
21 direct knowledge.

22 Q. (By Mr. Young) Would you be able to give

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 68

1 me the name of one person who would be able to tell
2 me, one way or the other, whether or not LibGen was
3 used to train LLaMA 3?

4 A. Not sure --

5 (Stenographer clarification.)

6 THE WITNESS: -- I have a name for you.

7 Q. (By Mr. Young) What about Books3? Would
8 you be able to give me a name of a person who would
9 be able to tell me whether Books3 was used to train
10 the LLaMA 2 series of models?

11 A. I don't recall the shape of the
12 organization when we were developing LLaMA 3. So
13 I'm not sure who would, at that time, have the best
14 knowledge, most direct knowledge of that.

15 Q. And, I'm sorry, Mr. Al-Dahle. Did you mean
16 LLaMA 2 or LLaMA 3?

17 A. You asked LLaMA 3.

18 Q. I asked about LLaMA 2. So I apologize, if
19 you were answering about LLaMA 3, I will -- let me
20 just clean that up a little bit.

21 So, Mr. Al-Dahle, would you be able to give
22 me the name of a person who would be able to tell me

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 69

1 whether or not Books3 was used to train the LLaMA 2
2 series of models?

3 A. I don't know who would directly know. I'm
4 not sure who was working on datasets at those times.
5 I can't recall the person.

6 Q. What about LLaMA 3? Would you be able to
7 give me the name of a person who would be able to
8 tell me, one way or another, if Books3 was used to
9 train LLaMA 3?

10 A. I don't know definitively who would know.
11 So anything I give would be a guess.

12 Q. But is it your testimony today that you,
13 yourself, at least sitting right here, are not aware
14 of whether Books3 or LibGen were used to train
15 either LLaMA 2 or LLaMA 3; correct?

16 MR. GHAJAR: Objection. Vague.

17 THE WITNESS: I feel like I answered that.
18 The -- I was made aware of it recently as part of
19 the discussions with my lawyers.

20 MR. YOUNG: And I would caution the witness
21 not to reveal any substance of any conversations
22 with his lawyers.

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 226

1 on a lot of repetition that's in the datasets. And
2 it wasn't -- we weren't building the
3 individualization tools so that we can understand
4 where we were having software bugs in those data
5 pre-processing pipelines. So it was an
6 implementation issue and execution issue in the
7 team.

Q. And that's why the --

9 (Stenographer clarification.)

10 Q. (By Mr. Young) -- datasets needed to be
11 audited; correct?

12 A. That's why you need to build data
13 visualization tools, to identify mistakes in your
14 pre-possessing pipelines.

15 Q. Can you go to your message of 2933, please,
16 and then can you please read that message -- that
17 message.

18 A. "Do we have the right datasets in there?"

19 Q. Can you please read the next message.

20 A. "Is there anything you wanted to use but
21 couldn't for some stupid reason?"

22 Q. And Mr. Touvron responds to you. Can you

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 227

1 please read Mr. Touvron's response.

2 A. "LibGen is a good one. I'm currently using
3 it."

4 Q. And then what did you say after that?

5 A. "Yep, I cleared the path to use that one."

6 Q. Do you remember what stupid reasons you
7 were referring to when you said, "Is there anything
8 you wanted to use but couldn't for some stupid
9 reason"?

10 A. I don't have specifics of that, but if
11 you're in my role, you'd be surprised how many times
12 people do reason -- or do things that they -- that
13 they shouldn't be doing. So there's collaboration
14 issues, there's technical implementation issues,
15 there's people who -- there's all sorts of reasons
16 that require my help at times.

17 So, you know, he -- you know, he could have
18 not been able to get -- he's a research scientist.
19 He may not have been able to get, you know, data
20 scientists or data engineers to collaborate with
21 him.

22 Q. But, specifically, you are referring to

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 228

1 whether or not Mr. Touvron had the right datasets
2 that he couldn't access for some, quote/unquote,
3 "stupid reason"; correct?

4 A. In that context, I was looking for -- it's
5 unclear what I was looking for, but the context is
6 that, in my role, there's -- people do all kinds of
7 things that I need to look into.

8 Q. Now, LibGen, we talked about LibGen earlier
9 today; right?

10 A. Yes.

11 Q. Do you remember what you testified to me
12 about, whether or not you had heard of LibGen
13 before?

14 A. Yeah, I don't have a specific memory of it.

15 Q. Do you recall testifying earlier today that
16 you heard about LibGen about a month or so before?

17 A. Yes.

18 MR. GHAJAR: Objection. Vague.

19 Q. (By Mr. Young) Do you stand by that
20 testimony, that the first time you heard about
21 LibGen was about a month or -- month or so before
22 today?

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 229

1 A. That wasn't my testimony. I said I don't
2 have a specific memory of discussing LibGen before.

3 Q. Now, based on what you just read here, and
4 based on the messages that you and Mr. Touvron
5 shared, is it your understanding that you had an
6 awareness of LibGen at least as early as October
7 12th, 2023?

8 A. I guess it was mentioned to me in a chat
9 amongst, likely, thousands of chats at the time.

10 Q. You said you, quote/unquote, "cleared the
11 path to use that one"; right?

12 A. Yes.

13 Q. What does that mean?

14 A. I'm not sure. Could have been the process
15 I'm referring to earlier, where we needed to do a
16 review with a set of experts across the company and
17 have a discussion on it.

18 Q. That's right. I've heard that answer
19 before. You do review with a variety of experts
20 which could include lawyers, policy; right?

21 A. That's right.

22 Q. Does that mean -- does "clear the path"

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 230

1 mean override that process?

2 A. No.

3 Q. Okay.

4 Then what does it mean?

5 A. I don't have a specific recollection of
6 this, but it could have been something as simple as
7 initiating the process.8 Q. Were you aware of experiments that Meta was
9 running involving LibGen database or generative AI
10 models before October of 2023?

11 A. No.

12 Q. For example, were you aware that in
13 December of 2022, Mr. Aurelian Rodriguez had ran
14 experiments involving LibGen and Sci-Hub for the --
15 for Project Genesis? Were you aware of that?

16 MR. GHAJAR: Objection. Assumes facts.

17 THE WITNESS: Who is the -- sorry. Who did
18 you just reference?19 Q. (By Mr. Young) Do you know who
20 Aurelian Rodriguez is?21 A. The name is vaguely familiar, but I can't
22 remember who they are.

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 231

1 Q. Would it surprise you that Mr. Rodriguez
2 and other members of the Project Genesis team had
3 ran experiments on LibGen as early as December 2022?

4 MR. GHAJAR: Objection. Assumes facts.

5 THE WITNESS: I'll remind you that I wasn't
6 part of the core team for Genesis, so I don't have
7 that kind of visibility or understanding.

8 Q. (By Mr. Young) Were you aware that in
9 December of 2022, Dr. Presani had expressed concerns
10 about the use of LibGen in any published models that
11 Meta was to distribute?

12 MR. GHAJAR: Objection. Assumes facts.
13 Lacks foundation.

14 THE WITNESS: I don't know who Dr. --
15 sorry. What was the name again?

16 Q. (By Mr. Young) Presani.

17 A. I don't -- I'm not familiar with him.

18 MR. GHAJAR: Mr. Young, we've been going
19 about an hour and 10, approaching an hour and 15
20 minutes.

21 MR. YOUNG: I'll be wrapping this document
22 up in a couple of minutes and then we can break.

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 232

1 MR. GHAJAR: Great.

2 Q. (By Mr. Young) Now, do you see your
3 message of 3050 PDT?4 A. 3050, "I can clear the path for other
5 things."

6 Q. That's right.

7 What other things were you referring to?

8 A. I'm not sure, but I was soliciting feedback
9 on what's slowing him down.10 Q. Were you inviting Mr. Touvron to suggest
11 other obstacles you could clear for his job?12 A. I was suggesting he provide suggestions so
13 that I could have a better understanding of how the
14 collaboration and the work is happening on the
15 ground, and if I can do anything to help or initiate
16 a process to help.17 Q. For example, clearing the path to use
18 LibGen?19 A. Which would have involved me initiating the
20 process.21 Q. Now, you go on. You say, "We should";
22 right?

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 233

1 A. Yes.

2 Q. What do you say after that?

3 A. "This is a lot of fun. We're going to make
4 this awesome together. Me pushing from the top and
5 you making it happen."

6 Q. Do you see your message of 33 --

7 (Stenographer clarification.)

8 Q. (By Mr. Young) -- 3313 PDT?

9 A. "LLaMA 3 is literally all I care about."

10 Q. And at this time, Mr. Touvron was working
11 on developing LLaMA 3; correct?12 A. Yes, but there's context here. When
13 discussing it with Hugo, Hugo is the research
14 scientist working on that project. So I'm
15 reassuring and reaffirming my commitment to that
16 project, and making sure that he understands his
17 priorities as well. He has competing set of
18 projects.19 Q. And that includes making sure that Hugo has
20 the right datasets to train LLaMA 3; correct?21 A. It's making sure that Hugo is focused
22 explicitly on LLaMA 3 and not other side projects.

10/3/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Ahmad Al-Dahle

Page 286

1 I, HEATHER J. BAUTISTA, CSR No. 11600,
2 Certified Shorthand Reporter, certify:
3 That the foregoing proceedings were taken
4 before me at the time and place therein set forth,
5 at which time the witness declared under penalty of
6 perjury; that the testimony of the witness and all
7 objections made at the time of the examination were
8 recorded stenographically by me and were thereafter
9 transcribed under my direction and supervision; that
10 the foregoing is a full, true, and correct
11 transcript of my shorthand notes so taken and of the
12 testimony so given;

13 (XX) Reading and signing was not requested/offered.

14 I further certify that I am not financially
15 interested in the action, and I am not a relative or
16 employee of any attorney of the parties, nor of any
17 of the parties.

18 I declare under penalty of perjury under the
laws of California that the foregoing is true and
19 correct. Dated: October 4, 2024

20

21

22



HEATHER J. BAUTISTA, CSR, CRR, RPR, CLR